An Extension of a Factorization Theorem of Wedderburn to Matrix Rings

Thomas J. Laffey and Eleanor Meehan
Department of Mathematics
University College Dublin
Dublin 4, Ireland

Submitted by Biswa Datta

ABSTRACT

Let F be a field of characteristic zero, and let $M_n(F)$ be the algebra of $n \times n$ matrices over F. Let $f(x)$ be a monic polynomial of degree n in $F[x]$. It is proved that there exist $n \times n$ matrices $A_1, \ldots, A_n$ all with minimal polynomial $f(x)$ such that

$$f(x)I_n = (xI_n - A_1) \cdots (xI_n - A_n).$$

A simple inductive procedure for constructing $A_2, \ldots, A_n$, having chosen A_1 to be the companion matrix of $f(x)$, is established. The procedure also leads to an improved version of a theorem of Wedderburn on the factorization of certain polynomials over division rings.

INTRODUCTION

Let D be a division algebra finite dimensional over its center F, and let $f(x) \in F[x]$ be a monic irreducible polynomial of degree n. Then a theorem of Wedderburn [1, p. 72; 6, p. 179] states that if there exists an element $d \in D$ with $f(d) = 0$, then $f(x)$ has a factorization

$$f(x) = (x - d_1)(x - d_2) \cdots (x - d_n)$$

where $d_1 = d, d_2, \ldots, d_n$ are all conjugate to d in D. In this paper we present an explicit version of this theorem and extend the result to matrix rings.

655 Avenue of the Americas, New York, NY 10010 0024-3795/92/$5.00

1. PRELIMINARIES

Let R be a ring with identity, and let $R[x]$ be the ring of all polynomials with coefficients in R. For $f(x)$, $g(x) \in R[x]$, we say $f(x)$ divides $g(x)$ if $g(x) = f(x)h(x)$ for some $h(x) \in R[x]$. Note that $f(x)$ is on the left here. Since we have not assumed R is commutative, the definition is not left-right symmetric.

If $f(x) = a_0 + xa_1 + \cdots + x^r a_r \in R[x]$ and $z \in R$, we write $f(z) = a_0 + za_1 + \cdots + z^r a_r$. (Note the substitution on the left of the coefficients.) Note that

$$x^k - z^k = (x - z)(x^{k-1} + x^{k-2}z + \cdots + z^{k-1})$$

and thus

$$x - z \quad \text{divides} \quad f(x) - f(z).$$

Suppose now that $f(x) = g(x)h(x)$ where $f(x)$, $g(x)$, $h(x) \in R[x]$. If for some $a \in R$, $f(a) = 0$ and $g(a)$ is a unit of R, then from the equations

$$f(x) = f(x) - f(a) = (x - a)t(x)$$

and

$$\begin{aligned} f(x) &= [g(x) - g(a)]h(x) + g(a)h(x) \\ &= (x - a)s(x) + g(a)h(x) \end{aligned}$$

we obtain

$$h(x) = (x - a^{g(a)})h_1(x),$$

where we use the notation a^w for $w^{-1}aw$ if w is a unit. This yields

$$f(x) = g_1(x)h_1(x),$$

where

$$g_1(x) = g(x)(x - a^{g(a)}).$$

Suppose $f(x) \in R[x]$ is a monic polynomial. The above procedure can be carried out inductively if $f(a) = 0$ for a set of elements $a \in R$ for which at each stage the appropriate $g(a)$ is a unit. Of course, if R is a division algebra, the requirement that $g(a)$ be a unit is equivalent to the condition that $g(a) \neq 0$. This describes the method of Wedderburn (see [6, pp. 178–182] for further information). In the division ring situation, Wedderburn argued by contradiction to show that sufficient conjugates exist to enable the factorization to be carried out (if $f(x) \in F[x]$ is irreducible and $f(a) = 0$ for some $a \in D$) and to express $f(x)$ as the product of linear factors. In Section 3 here we obtain a more explicit version of his result.

2. THE FACTORIZATION FORMULA

Suppose R is a ring with identity, $Z(R)$ its center, and $f(x) \in Z(R)[x]$ a monic polynomial of degree n. Suppose $f(a) = 0$ for some $a \in R$. Write $a_1 = a$. Then

$$f(x) = (x - a_1)f_1(x)$$

for a monic $f_1(x) \in R[x]$. Suppose $b \in R$ is a unit such that $a_1^b - a$ is a unit of R. Since $f(a_1^b) = [f(a_1)]^b = 0$, we have the situation described in Section 1 and we obtain

$$f(x) = (x - a_1)(x - a_2)f_2(x),$$

where $a_2 = a^{b(a_1^b - a_1)} = a_1^{c_1}$, where $c_1 = [a_1, b]$, where as usual $[p, q]$ denotes $pq - qp$. Having obtained $f(x) = u_k(x)f_k(x)$ where $u_k(x) = (x - a_1) \cdots (x - a_k)$ and $k < n$, we suppose that $u_k(a^{b^k})$ is a unit of R. By the method of Section 1, we then may write

$$f_k(x) = (x - a_{k+1})f_{k+1}(x),$$

where

$$a_{k+1} = a^{v_k(a^{b^k})} \quad \text{and} \quad v_k(a^{b^k}) = b^k u_k(a^{b^k}).$$

We now obtain a simple inductive formula for the sequence a_k and $v_k(a^{b^k})$. Suppose first that $k = 1$. Then

$$v_1(a^b) = [a, b] = c_1$$

and thus

$$a_2 = a_1^{c_1}.$$

Next,

$$\begin{aligned} v_2\left(a^{b^2}\right) &= a^2b^2 - ab^2\left(a_1 + a_2\right) + b^2a_1a_2 \\ &= a\left[a, b^2\right] - \left[a, b^2\right]a_2 \\ &= ac_1b + abc_1 - bc_1a_2 - c_1ba_2 \\ &= ac_1b + bac_1 + c_1^2 - bc_1a_2 - c_1ba_2 \\ &= c_1a_2b + bac_1 + c_1^2 - bac_1 - c_1ba_2, \end{aligned}$$

using $c_1a_2 = a_1c_1$. Thus

$$\begin{aligned} v_2\left(a^{b^2}\right) &= c_1^2 + c_1c_2 \qquad \left(\text{where} \quad c_2 = \left[a_2, b\right]\right) \\ &= c_1\left(c_1 + c_2\right). \end{aligned}$$

So $a_3 = a^{c_1(c_1+c_2)} = a_2^{c_1+c_2}$.

This leads to the following assertion, the proof of which occupies most of this section.

CLAIM 1. *For $k = 1, 2, \ldots, n-1$, one has $v_k(a^{b^k}) = d_1d_2 \cdots d_k$, where $d_1 = c_1 = [a, b]$, and for $i \geqslant 1$, $d_{i+1} = d_i + c_{i+1}$, where $c_i = [a_i, b]$.*

Before beginning the proof of the claim, we note that it yields a simple inductive definition of the sequence $a_1, a_2, \ldots, a_n$; namely: set $d_1 = c_1 = [a_1, b]$, $a_2 = a_1^{c_1}$, $c_2 = [a_2, b]$, $d_2 = d_1 + c_2$, $a_3 = a_2^{d_2}, \ldots, c_i = [a_i, b]$, $d_i = d_{i-1} + c_i$, $a_{i+1} = a_i^{d_i}$, $i = 1, 2, \ldots, n-1$.

The calculation given above yields the claim for $k = 1$ and $k = 2$. However, the general situation is much more complicated, and we prove the result indirectly by establishing a stronger formula for which it is easier to construct an inductive argument.

Extending the notation above, we set $v_k(a^{b^l}) = b^lu_k(a^{b^l})$ for $k = 1, 2, \ldots, n-1$ and all nonnegative integers l. In order to motivate the next claim, we note that

$$\begin{aligned} v_1\left(a^{b^k}\right) &= \left[a, b^k\right] \\ &= \sum_{l=0}^{k-1} b^lc_1b^{k-1-l}, \end{aligned}$$

so $v_1(a^{b^k})$ is the sum of all elements of the form $b^{j_1}c_1b^{j_2}$ where j_1, j_2 are nonnegative integers subject to the condition $j_1 + j_2 = k - 1$. We now assert

CLAIM 2. *For* $m = 1, 2, \ldots, n-1$,

$$v_m\left(a^{b^k}\right) = 0 \qquad \text{if} \quad 1 \leqslant k < m$$

and

$$v_m\left(a^{b^k}\right) = \text{sum of all elements of the form}$$

$$b^{j_1}d_1b^{j_2}d_2 \cdots b^{j_m}d_mb^{j_{m+1}}$$

where $j_1, j_2, \ldots, j_{m+1}$ *are nonnegative integers*

$$\text{with } \left(j_1 + j_2 + \cdots + j_{m+1} = k - m\right) \qquad \text{if} \quad k \geqslant m.$$

(*Here* $c_i = [a_i, b]$ *and* $d_i = c_1 + c_2 + \cdots + c_i$ *as in Claim* 1.)

We prove the claim by induction on m (for all positive integers k), the case $m = 1$ having been established above.

Suppose $m > 1$ and the claim holds for $v_1, \ldots, v_{m-1}$. Suppose it also holds for $v_m(a^{b^l})$ with $1 \leqslant l < k$. Note that Claim 1 then holds for $v_1, \ldots, v_{m-1}$, and thus in particular $a_1, \ldots, a_m$ satisfy the inductive relations $a_{i+1} = a_i^{d_i}$ for $i \leqslant m - 1$. Now

$$\begin{aligned} v_m\left(a^{b^k}\right) &= a^m b^k - a^{m-1}b^k\left(a_1 + \cdots + a_m\right) \\ &\quad + a^{m-2}b^k\left(a_1a_2 + a_1a_3 + \cdots + a_1a_m + a_2a_3 + \cdots + a_{m-1}a_m\right) \\ &\quad - \cdots + \left(-1\right)^m b^k a_1a_2 \cdots a_m \\ &= av_{m-1}\left(a^{b^k}\right) - v_{m-1}\left(a^{b^k}\right)a_m. \end{aligned}$$

If $k < m - 1$, then by induction $v_{m-1}(a^{b^k}) = 0$ and thus $v_m(a^{b^k}) = 0$ as required. If $k = m - 1$, then

$$av_{m-1}\left(a^{b^{m-1}}\right) = v_m\left(a^{b^{m-1}}\right)a_m$$

by the definition of a_m, so again $v_m(a^{b^k}) = 0$. Hence we may assume $k \geqslant m$. To make the general argument a little easier to follow, we first do the case $k = m$. Now

$$v_m\left(a^{b^m}\right) = av_{m-1}\left(a^{b^m}\right) - v_{m-1}\left(a^{b^m}\right)a_m.$$

By induction $v_{m-1}(a^{b^m})$ is the sum of all terms of the form

$$bd_1d_2 \cdots d_{m-1}, d_1bd_2 \cdots d_{m-1}, \ldots, d_1d_2 \cdots d_{m-1}b.$$

Note that

$$abd_1d_2 \cdots d_{m-1} = bad_1d_2 \cdots d_{m-1} + c_1d_1d_2 \cdots d_m,$$

and using the relations $a_id_i = d_ia_{i+1}$ for $i = 1, 2, \ldots, m-1$,

$$bad_1d_2 \cdots d_{m-1} = bd_1d_2 \cdots d_{m-1}a_m.$$

Hence

$$a\left(bd_1d_2 \cdots d_{m-1}\right) - \left(bd_1d_2 \cdots d_{m-1}\right)a_m = c_1d_1d_2 \cdots d_{m-1}.$$

Similarly

$$\begin{aligned} &ad_1bd_2 \cdots d_{m-1} - d_1bd_2 \cdots d_{m-1}a_m \\ &\quad = d_1a_2bd_2 \cdots d_{m-1} - d_1bd_2 \cdots d_{m-1}a_m \\ &\quad = d_1c_2d_2 \cdots d_{m-1} + d_1ba_2d_2 \cdots d_{m-1} - d_1bd_2 \cdots d_{m-1}a_m \\ &\quad = d_1c_2d_2 \cdots d_{m-1}, \end{aligned}$$

and in general

$$\begin{aligned} &ad_1d_2 \cdots d_ibd_{i+1} \cdots d_{m-1} - d_1d_2 \cdots d_ibd_{i+1} \cdots d_{m-1}a_m \\ &\quad = d_1d_2 \cdots d_ic_{i+1}d_{i+1} \cdots d_{m-1} \qquad \text{for} \quad i = 1, 2, \ldots, m-2 \end{aligned}$$

and

$$ad_1d_2 \cdots d_{m-1}b - d_1d_2 \cdots d_{m-1}ba_m = d_1d_2 \cdots d_{m-1}c_m.$$

Hence

$$v_m(a^{b^m}) = c_1d_1d_2 \cdots d_{m-1} + d_1c_2d_2 \cdots d_{m-1} + d_1d_2c_3d_3 \cdots d_{m-1} + \cdots + d_1d_2 \cdots d_{m-1}c_m.$$

Using the fact that $d_i = c_1 + c_2 + \cdots + c_i = d_{i-1} + c_i$ for $i = 1, 2, \ldots, m - 1$, a simple inductive argument now yields that

$$v_m(a^{b^m}) = d_1d_2 \cdots d_{m-1}d_m,$$

where $d_m = c_1 + c_2 + \cdots + c_m$.

Suppose $k > m$. Using the induction hypothesis for $v_{m-1}(a^{b^k})$, we consider a typical term

$$w = b^{j_1}d_1b^{j_2}d_2 \cdots b^{j_{m-1}}d_{m-1}b^{j_m}$$

in its expansion and look at $aw - wa_m$. Now

$$\begin{aligned}
&ab^{j_1}d_1b^{j_2}d_2 \cdots d_{m-1}b^{j_m} \\
&\quad = b^{j_1}ad_1b^{j_2}d_2 \cdots d_{m-1}b^{j_m} + \left[a_1, b^{j_1}\right]d_1b^{j_2} \cdots d_{m-1}b^{j_m} \\
&\quad = b^{j_1}d_1b^{j_2}a_2d_2 \cdots d_{m-1}b^{j_m} + \left[a, b^{j_1}\right]d_1b^{j_2} \cdots d_{m-1}b^{j_m} \\
&\qquad + b^{j_1}d_1\left[a, b^{j_2}\right]d_2 \cdots d_{m-1}b^{j_m} \\
&\quad = \left[a, b^{j_1}\right]d_1b^{j_2} \cdots d_{m-1}b^{j_m} + b^{j_1}d_1\left[a_2, b^{j_2}\right]d_2 \cdots d_{m-1}b^{j_m} \\
&\qquad + \cdots + b^{j_1}d_1b^{j_2}d_2 \cdots b^{j_{m-2}}d_{m-2}\left[a_{m-1}, b^{j_{m-1}}\right]d_{m-1}b^{j_m} \\
&\qquad + b^{j_1}d_1b^{j_2}d_2 \cdots b^{j_{m-1}}d_{m-1}\left[a_m, b^{j_m}\right] \\
&\qquad + b^{j_1}d_1b^{j_2}d_2 \cdots b^{j_{m-1}}d_{m-1}b^{j_m}a_m.
\end{aligned}$$

Using the equation

$$\begin{aligned}
\left[a_i, b^r\right] &= \sum_{t=0}^{r-1} b^t\left[a_i, b\right]b^{r-t-1} \\
&= \sum b^tc_ib^{r-t-1},
\end{aligned}$$

we see that

$$ab^{j_1}d_1b^{j_2}d_2 \cdots b^{j_m}d_{m-1}b^{j_m} - b^{j_1}d_1b^{j_2}d_2 \cdots d_{m-1}b^{j_m}a_m$$

is the sum of all terms (if any) of the following forms:

$$b^{r_1}c_1b^{r_2}d_1b^{j_2} \cdots d_{m-1}b^{j_m}, \qquad r_1 \geqslant 0, \quad r_2 \geqslant 0, \quad r_1 + r_2 = j_1 - 1,$$

$$b^{j_1}d_1b^{s_1}c_2b^{j_2}d_2b^{j_3} \cdots d_{m-1}b^{j_m}, \qquad s_1 \geqslant 0, \quad s_2 \geqslant 0, \quad s_1 + s_2 = j_2 - 1,$$

$$\vdots$$

$$b^{j_1}d_1b^{j_2}d_2 \cdots b^{j_{m-1}}d_{m-1}b^{t_1}c_mb^{t_2}, \qquad t_1 \geqslant 0, \quad t_2 \geqslant 0,$$
$$t_1 + t_2 = j_m - 1.$$

Again using the fact that the sum of the terms

$$c_1d_1d_2 \cdots d_{m-1}, \qquad d_1c_2d_2d_3 \cdots d_{m-1}, \qquad d_1d_2c_3d_3 \cdots d_{m-1}, \ldots,$$

$$d_1d_2 \cdots d_{m-1}c_m$$

equals $d_1d_2 \cdots d_m$ and noting that the above sum represents all terms obtained by inserting an extra factor c_i into the product $b^{j_1}d_1b^{j_2}d_2 \cdots b^{j_{m-1}}d_{m-1}b^{j_m}$, it follows that as $j_1, \ldots, j_m$ varies over all nonnegative integers with $j_1 + \cdots + j_m = k - m + 1$ we obtain exactly the sum of the terms

$$b^{l_0}c_1b^{l_1}d_1b^{l_2}d_2 \cdots d_{m-1}b^{l_m},$$

$$b^{l_0}d_1b^{l_1}c_2b^{l_2}d_2 \cdots d_{m-1}b^{l_m},$$

$$\vdots$$

$$b^{l_0}d_1b^{l_1}d_2 \cdots b^{l_{m-2}}d_{m-1}b^{l_{m-1}}c_mb^{l_m},$$

over all nonnegative integers $l_0, \ldots, l_m$ with $l_0 + \cdots + l_m = k - m$ and that this equals the sum of all such terms

$$b^{l_0}d_1b^{l_1}d_2 \cdots b^{l_{m-1}}d_mb^{l_m},$$

as required.

This establishes Claim 2 and also Claim 1.

We summarize the results as follows.

THEOREM 1. *Let R be a ring with identity, and $f(x) \in Z(R)[x]$ a monic polynomial of degree n. Suppose that $f(a) = 0$ for some $a \in R$ and that there exists $b \in R$ for which $d_1, \ldots, d_{n-1}$ defined inductively below are units of R. Then*

$$f(x) = (x - a_1) \cdots (x - a_n)$$

where $a_1 = a$, $a_{i+1} = a_i^{d_i}$ $(i = 1, 2, \ldots, n-1)$.

The d_i are defined as follows: $a_1 = a$, $d_1 = c_1 = [a_1, b]$, $a_2 = a_1^{d_1}$, $c_2 = [a_2, b]$, $d_2 = c_1 + c_2, \ldots,$ $a_{k+1} = a_k^{d_k}$, $c_{k+1} = [a_{k+1}, b]$, $d_{k+1} = d_k + c_{k+1}$.

3. THE DIVISION CASE

Suppose D is a finite dimensional division algebra over its center F, and let $f(x) \in F[x]$ be a monic irreducible polynomial of degree $n > 1$ over F. Suppose $f(a) = 0$ for some $a \in D$. We show that the procedure described in Theorem 1 can be carried out to yield a factorization of $f(x)$. The procedure depends on being able to choose $b \in D$ in such a way that the elements $d_1, d_2, \ldots, d_{n-1}$ defined in Theorem 1 are nonzero. Since $n > 1$ and $f(x)$ is irreducible in $F[x]$, a is not in F, so we can choose $b \in D$ with $ab \neq ba$, yielding $d_1 \neq 0$. Suppose $b \in D$ has been chosen so that $d_1, d_2, \ldots, d_{k-1}$ are nonzero and that d_k is zero. The equation $d_k = 0$ means that

$$a^k b^k - a^{k-1} b^k (a_1 + \cdots + a_k) + a^{k-2} b^k \left(\sum_{1 \leqslant i < j \leqslant k} a_i a_j \right) - \cdots$$

$$+ (-1)^k b^k a_1 a_2 \cdots a_k = 0.$$

This may be written in the form $aw = wa_k$, where

$$w = a^{k-1} b^k - a^{k-2} b^k (a_1 + \cdots + a_{k-1})$$

$$+ a^{k-2} b^k \left(\sum_{1 \leqslant i < j \leqslant k-1} a_i a_j \right) - \cdots + (-1)^{k-1} b a_1 a_2 \cdots a_{k-1}.$$

By definition

$$a_k = v^{-1} a v,$$

where

$$v = a^{k-1}b^{k-1} - a^{k-2}b^{k-1}(a_1 + \cdots + a_{k-1}) + \cdots .$$

Hence $w = x_1 v$, where x_1 is an element satisfying $x_1 a = a x_1$. Thus we may write

$$\begin{aligned} a\Big(& a^{k-2}\big(b^k - x_1 b^{k-1}\big) - a^{k-3}\big(b^k - x_1 b^{k-1}\big)(a_1 + \cdots + a_{k-2}) \\ & + a^{k-4}\big(b^k - x_1 b^{k-1}\big) \sum_{1 \leqslant i < j \leqslant k-2} a_i a_j - \cdots \Big) \\ = & \Big[a^{k-2}\big(b^k - x_1 b^{k-1}\big) - a^{k-3}\big(b^k - x_1 b^{k-1}\big) \\ & \times (a_1 + \cdots + a_{k-2}) + \cdots \Big] a_{k-1}. \end{aligned}$$

Since $az = za_{k-1}$, where

$$\begin{aligned} z = & a^{k-2}b^{k-2} - a^{k-3}b^{k-2}(a_1 + \cdots + a_{k-2}) + \cdots \\ & + (-1)^{k-2} b^{k-2} a_1 a_2 \cdots a_{k-2}, \end{aligned}$$

we find that

$$\begin{aligned} & a^{k-2}\big(b^k - x_1 b^{k-1}\big) - a^{k-3}\big(b^k - x_1 b^{k-1}\big)(a_1 + \cdots + a_{k-2}) \\ & \quad + \cdots + \big(b^k - x_1 b^{k-1}\big)(a_1 a_2 \cdots a_{k-2}) \\ & = x_2 \Big[a^{k-2}b^{k-2} - a^{k-3}b^{k-2}(a_1 + \cdots + a_{k-2}) \\ & \qquad + \cdots + (-1)^{k-2} b^{k-2} a_1 a_2 \cdots a_{k-2} \Big] \end{aligned}$$

for some x_2 with $ax_2 = x_2 a$. This yields

$$aq = q a_{k-2},$$

where

$$\begin{aligned} q = & a^{k-3}\big(b^k - x_1 b^{k-1} - x_2 b^{k-2}\big) - a^{k-4}\big(b^k - x_1 b^{k-1} - x_2 b^k\big) \\ & \cdot (a_1 + \cdots + a_{k-3}) + \cdots . \end{aligned}$$

Proceeding thus, we eventually find that there exist $x_1, x_2, \ldots, x_{k-1}$ in D with $x_i a = a x_i$ such that

$$a\left(b^k - x_1 b^{k-1} - x_2 b^{k-2} - \cdots - x_{k-1} b\right) = \left(b^k - x_1 b^{k-1} - x_2 b^{k-2} - \cdots - x_{k-1} b\right) a,$$

so

$$b^k = x_1 b^{k-1} + x_2 b^{k-2} + \cdots + x_{k-1} b + x_k$$

for some x_k with $a x_k = x_k a$.

Let $C = \{x \in D \mid xa = ax\}$. The last equation shows that if $d_k = 0$, then the vector space sum

$$Cb^k + Cb^{k-1} + \cdots + Cb + C$$

is not direct. We need to show that b can be chosen so that none of the elements $d_1, d_2, \ldots, d_{n-1}$ is zero; thus it suffices to show that b can be chosen in D so that the sum

$$Cb^{n-1} + Cb^{n-2} + \cdots + Cb + C$$

is direct.

By standard results of Jacobson [2, Chapter VII], $F(a)$ is the center of C, $\dim D = n \dim C$, and there is a maximal subfield M of D containing a with M separable over $F(a)$. One can then deduce that $M = F(t)$ for some element $t \in M$. Let $M = F(t)$. By Jacobson [2, Corollary to Theorem, p. 182] there exists $b \in D$ such that

$$\left\{t^i b^j \middle| i, j = 0, 1, \ldots, n-1\right\}$$

is a vector space basis of D over F, where $n^2 = \dim D$. In particular D is the vector space sum $C + Cb + \cdots + Cb^{n-1}$. If the sum $C + Cb + \cdots + Cb^k$ is not direct, then viewing it as a left C-vector space, we find that there exists $j \leqslant k$ with

$$Cb^j \subseteq C + Cb + \cdots + Cb^{j-1}.$$

But then $Cb^l \subseteq C + Cb + \cdots + Cb^{j-1}$ for all l, and $\dim D \leqslant (j-1) \dim C < k \dim C = \dim D$, which is false. Hence the sum

$$C + Cb + \cdots + Cb^{k-1}$$

is direct, as required.

REMARK. Other "natural" conjugates of a than a^b, a^{b^2}, a^{b^3}, ... may be used in carrying out the Wedderburn factorization process to give different choices for $a_2, a_3, \ldots$. These yield interesting formulae also for $n = 3$ and $n = 4$. See Rowen [6, pp. 180–182] for details. We have not been able to get a straightforward inductive formula for these other choices for general n.

4. MATRIX RINGS OVER FIELDS

Let F be a field, and $f(x) \in F[x]$ a monic polynomial of degree n. We now discuss the problem of factorizing $f(x)I_n$ in the form

$$(xI_n - A_1)(xI_n - A_2) \cdots (xI_n - A_n)$$

where $A_1, A_2, \ldots, A_n$ are nonderogatory matrices in $M_n(F)$ with characteristic polynomial $f(x)$. In carrying out the Wedderburn process starting (without loss of generality) with A_1 being the companion matrix of $f(x)$, we must find a conjugate A_2 $(= T^{-1}AT)$ such that $A_2 - A_1$ is nonsingular. At the next stage, we must find a conjugate A_3 with

$$A_3^2 - A_3(A_1 + A_2) + A_1A_2$$

nonsingular, etc. One approach will be to use multilinear algebra techniques [5, Chapter 7]. Writing $A_2 = T^{-1}AT$ etc., one is naturally led to consider the image of the map

$$A \otimes I - I \otimes A^T : M_n(F) \to M_n(F)$$
$$: X \to AX - XA$$

in the first case,

$$(A \otimes I - I \otimes A^T)(A \otimes I - I \otimes A_2^T)$$

in the second case, and so on, the aim in each case being to prove that the image under consideration contains a nonsingular matrix. Over an algebraically closed field, it should be possible to use algebraic-geometric methods to do this, though the singularities of the mapping cause difficulties. At the last stage of the construction, the image in question is just n-dimensional, so a purely vector-space argument would not appear to have much likelihood of success. The arguments given here work with the formula of Section 2 directly

and do not require the algebraic closure of F. The discussion does require the field to be infinite or have characteristic greater than n. In order to make the exposition clearer, we assume here that *F has characteristic zero.* It is not difficult to modify the method for infinite fields of prime characteristic, but the formulae become more unwieldy. The case of finite fields will be treated elsewhere [3].

Suppose then F is a field of characteristic zero, $f(x) \in F[x]$ a monic (not necessarily irreducible) polynomial of degree n. Let A_1 be the modified companion matrix of $f(x)$ obtained by performing a similarity on the ordinary companion matrix $C(f)$ using the diagonal matrix

$$\left[\operatorname{diag}\big((n-1)!, (n-2)!, \ldots, 2!, 1!, 1\big)\right]^{-1}.$$

So A is of the form

$$\begin{bmatrix} 0 & n-1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & n-2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{n,n-1} & a_{nn} \end{bmatrix}.$$

Let

$$B = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 1 \end{bmatrix}.$$

We now show that the process of Section 2 can be carried out here. Let $A_1 = A$. Note first that the commutator $C_1 = [A, B]$ is lower triangular with diagonal

$$\operatorname{diag}(C_1) = (n-1, -1, -1, \ldots, -1),$$

so C_1 is invertible, and its inverse is lower triangular with diagonal

$$\left(\frac{1}{n-1}, -1, -1, \ldots, -1\right).$$

We set $A_2 = C_1^{-1}A_1C_1$. Note that A_2 is of the form

$$\begin{bmatrix} y_{11} & -1 & 0 & 0 & 0 & \cdots & 0 \\ y_{21} & y_{22} & n-2 & 0 & 0 & \cdots & 0 \\ y_{31} & y_{32} & y_{33} & n-3 & 0 & \cdots & 0 \\ \cdot & & & \cdot & \cdot & \cdot & \vdots \\ \cdot & & & & \cdot & 2 & 0 \\ \cdot & & & & & \cdot & 1 \\ y_{n1} & \cdot & \cdot & \cdot & \cdot & \cdot & y_{nn} \end{bmatrix},$$

so A_2 is lower Hessenberg with the first superdiagonal equal to $(-1, n-2, n-3, \ldots, 1)$. We now set $C_2 = [A_2, B]$. Then C_2 is again lower triangular with

$$\operatorname{diag}(C_2) = (-1, n-1, -1, \ldots, -1),$$

and thus, setting $D_2 = C_1 + C_2$, we find D_2 is lower triangular with

$$\operatorname{diag}(D_2) = (n-2, n-2, -2, \ldots, -2).$$

Then $A_3 = D_2^{-1}A_2D_2$ is lower Hessenberg with first superdiagonal equal to

$$(-1, -2, n-3, n-4, \ldots, 2, 1).$$

Proceeding by induction, suppose $k < n-1$ and A_k is lower Hessenberg with first superdiagonal equal to

$$(-1, -2, \ldots, -(k-1), n-k, n-k-1, \ldots, 2, 1),$$

and D_{k-1} is lower-triangular with

$$\operatorname{diag}(D_{k-1}) = \Big(\underbrace{n-k-1, n-k-1, \ldots, n-k-1}_{k-1}, \underbrace{-(k-1), \ldots, -(k-1)}_{n-(k-1)}\Big)$$

Then $C_k = [A_k, B]$ is lower triangular with

$$\operatorname{diag}(C_k) = (-1, -1, \ldots, -1, \underset{\substack{\uparrow \\ k\text{th position}}}{n-1}, -1, \ldots, -1).$$

So $D_k = D_{k-1} + C_k$ is lower triangular with

$$\operatorname{diag}(D_k) = (n-k, \ldots, n-k, n-k, -k, \ldots, -k),$$

and thus $A_{k+1} = D_k^{-1} A_k D_k$ is lower Hessenberg with

$$\operatorname{diag}(A_{k+1}) = (-1, -2, \ldots, -(k-1), -k, n-k-1, \ldots, 2, 1).$$

This establishes the inductive step. By Theorem 1 we thus have

$$f(x) I_n = (xI_n - A_1) \cdots (xI_n - A_n).$$

We record the result formally as

THEOREM 2. *Let F be a field of characteristic zero and $f(x) \in F[x]$ a monic polynomial of degree n. Then there exist $A_1, A_2, \ldots, A_n \in M_n(F)$ with each A_i nonderogatory and having characteristic polynomial $f(x)$ such that*

$$f(x) I_n = (xI_n - A_1) \cdots (xI_n - A_n).$$

EXAMPLE 1. $f(x) = x^3 - 2$. Then

$$(x^3 - 2) I_3 = (xI_3 - A_1)(xI_3 - A_2)(xI_3 - A_3),$$

where

$$A_1 = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ -2 & 0 & 0 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -2 \\ -1 & 0 & 0 \end{pmatrix}.$$

Note that A_1, A_2, A_3 do not commute here. Since the Galois group of $x^3 - 2$ over the rational field $\mathbf{Q}$ has order 6 and not 3, one cannot find commuting

3×3 matrices A_1, A_2, A_3 with rational entries satisfying

$$(x^3 - 2)I_3 = (xI_3 - A_1)(xI_3 - A_2)(xI_3 - A_3).$$

The question of "commuting factorization" in general will be discussed elsewhere [3].

EXAMPLE 2. Suppose $f(x) = x^n$. Here we take A_1 to be the matrix

$$\begin{bmatrix} 0 & n-1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & n-2 & 0 & \cdots & 0 \\ \cdot & & \ddots & \ddots & \ddots & \vdots \\ \cdot & & & 0 & 2 & 0 \\ \cdot & & & & 0 & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}.$$

It is easy to see that each D_k is diagonal in this case; in fact

$$D_k = \operatorname{diag}\left(\underbrace{n-k, n-k, \ldots, n-k,}_{k} -k, \ldots, -k\right)$$

and

$$A_{k+1} = \begin{bmatrix} 0 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & -2 & 0 & & & & & 0 \\ \vdots & & \ddots & \ddots & & & & & \vdots \\ 0 & \cdot & \cdot & 0 & -k & 0 & \cdot & \cdot & 0 \\ \cdot & & & & 0 & n-k-1 & 0 & \cdots & 0 \\ \cdot & & & & & \ddots & \ddots & \ddots & \vdots \\ \cdot & & & & & & 0 & 2 & 0 \\ \cdot & & & & & & & 0 & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}.$$

It may be worth remarking here that since the equation

$$x^n I_n = (xI_n - A_1) \cdots (xI_n - A_n)$$

involves only matrices with integer entries, given any prime number p, we can read it modulo p to yield a factorization of $x^n I_n$ valid over a field of characteristic p. However, if $p \leqslant n - 1$, at least one of the A_k has a 0 on its

first superdiagonal, and hence it is derogatory, and in particular it is not similar to the companion matrix of x^n. If F is a finite field, it may not be possible to factor $x^n I_n$ in the form $(xI_n - A_1) \cdots (xI_n - A_n)$ with each A_k upper triangular with minimal polynomial x^n, since each A_k must then have all the entries in its first superdiagonal nonzero. [In particular, if $F = \mathrm{GF}(2)$ and n is odd, $A_1 + \cdots + A_n$ must have first superdiagonal equal to $(1, 1, \ldots, 1)$, contrary to the requirement $A_1 + A_2 + \cdots + A_n = 0$.] Elsewhere [4], we present a factorization

$$x^n I_n = \left(xI_n - A_1\right) \cdots \left(xI_n - A_n\right)$$

with each A_k nonderogatory which is valid over every field.

In attempting an algebraic-geometric proof of Theorem 2, we may take A_1 to be the companion matrix of $f(x)$ and B to be a matrix of commuting indeterminates. Then we would define A_i, C_i, D_i inductively as in Theorem 1 and aim to prove that at each stage, $\det D_k$ is not identically zero. Then finally we would specialize the indeterminates to elements of F while preserving

$$\left(\det D_1\right) \cdots \left(\det D_{n-1}\right) \neq 0.$$

The difficulty in carrying out this process arises from the complexity of the expression for $\det D_k$ in terms of the entries of B. However, given the validity of Theorem 2, we know that the process will work. In numerical calculations, randomly generating a matrix B and using the inductive definitions of Theorem 1 to get $A_2, \ldots, A_n$ works very well, and the desired factorization is easy to achieve.

We note that the factorization yields information on the rank of certain commutators. For example, if $n \geqslant 3$, the nonsingularity of D_1, D_2 implies that if A is a nonderogatory $n \times n$ matrix, then there exists an $n \times n$ matrix B such that $[A, B]$ and $\left[\left[AB\right]^{-1}\left[A^2, B\right], B\right]$ are nonsingular. Equivalently, it can be used to obtain information on the image of the composition of certain linear mappings of the form $X \to AX - X(T^{-1}AT)$ from $M_n(F)$ to itself.

REFERENCES

1 I. N. Herstein, *Noncommutative Rings*, Carus Math. Monographs 15, Math. Assoc. Amer., 1968.

2 N. Jacobson, *Structure of Rings*, Amer. Math. Soc. Colloq. Pub. XXXVII, 1964.

3 T. J. Laffey and E. Meehan, Factorization of polynomials using commuting matrices, to appear.
4 T. J. Laffey and E. Meehan, Factorization of polynomials using unipotent Jordan blocks *Appl. Math. Lett.*, to appear.
5 M. Marcus, *Finite Dimensional Multilinear Algebra*, Marcel Dekker, New York, Vol. 1, 1973; Vol. 2, 1975.
6 L. H. Rowen, *Polynomial Identities in Ring Theory*, Academic, 1980.

Received 3 September 1991; final manuscript accepted 6 February 1992